MPA Response: Google's general search services - Proposed Conduct Requirements (Publishers)

February 2026

Summary and Recommendations

The UK music industry is a significant export sector. However, the UK market structure differs fundamentally from the US. While the US publishing market is largely domestic, the UK is an export powerhouse, with 43.8% of our revenue generated in foreign markets (compared to just 21.8% for the US).

This makes the UK uniquely vulnerable. If we erode our domestic copyright framework by accepting opaque training practices, we undermine our moral authority to enforce rights globally, placing approximately £745 million of annual export revenue at risk.

To protect this export value, we must address jurisdictional enforcement. UK copyright law must apply to any AI service operating in the UK, regardless of where the training took place. This rejects jurisdictional arbitrage, where firms scrape content abroad (in jurisdictions with weaker IP laws) to bypass UK law, but then deploy their models for commercial gain in the UK. We cannot allow SMS firms to import the benefits of infringement while outsourcing the liability.

The MPA recommends that the CMA amends the Publisher Conduct Requirement to:

1. **Re-evaluate proportionality:** Weigh the £150m cost to Google against the £408m risk to UK creators in the music publishing sector alone, and the 60% risk to the production music sector. It is notable that Google is budgeting to spend $185b on hardware and infrastructure to support its AI developments in 2026.
2. **Expand transparency**: Mandate the disclosure of training data sources to identify data laundering and copyright use, and that all and any AI-generated or AI-assisted content is clearly labelled at all times.
3. **Prioritise licensing:** Ensure technical controls do not supersede the legal requirement for valid licences.

 Introduction

The Music Publishers Association (MPA) is a membership organisation representing circa 180 UK music publishing companies, ranging from iconic independents to global major publishers. Via those members, the MPA represents over 95% of UK song rights. These rights are the foundation upon which the music industry is built.

The MPA owns the MCPS, the UK's mechanical copyright protecting and licensing society (representing over 42,000 direct composer, songwriter and music publisher members). The MPA also owns PMLL, which provides blanket licensing solutions whenever printed music of original compositions are photocopied within schools, higher education, choirs etc.

Using the analogy of a film and a script is a helpful way to understand what music publishing is, and how it fits into the wider music ecosystem: the recording of a song is the film itself and is owned by the record label. The song itself is the script and is owned by the songwriter and publisher. Music publishing is the business of managing the commercial life of that script—collecting royalties whenever it is used, whether it is remade, performed, or broadcast, regardless of which version of the film is playing.

The MPA represents this vital network of creators and businesses. Collectively, this sector generated £1.7 billion in 2024. This response is submitted on behalf of our members, who drive the vast majority of that economic contribution.

The impact of unlicensed AI training

Music publishing is, first and foremost, a sector that relies on licensing. It is a sector that has demonstrably shown itself to be adaptable to new technological developments—from radio to digital streaming—helping to create new markets and revenue streams that benefit both consumers and creators. We believe AI offers similar potential for innovation. However, this potential can only be realised if the government fosters a sustainable licensing environment.

Broad copyright exceptions that permit unlicensed AI training or use undermine the voluntary licensing solutions already emerging in the market. This creates a direct and serious financial risk to our members. The 2024 CISAC/PMP Strategy Global Study projects that 24% of creator revenues are at risk by 2028 if technology companies can process copyrighted works without a licence.

Unlike the live sector, our sector's value is derived almost entirely from copyright. Therefore, applying the CISAC projection to the annual music publishing revenue of £1.7 billion suggests a direct threat of losses starting at ~£408 million annually to songwriters and publishers based in the UK. This represents losses that the industry can ill afford, more so as they will only increase over time.

By artificially suppressing the cost of its primary input (training data) to zero, it not only inflicts severe economic harm on upstream creators but distorts the broader AI market. It creates insurmountable barriers to entry for competing AI developers who engage in fair, market-based licensing, thereby entrenching the dominance of Google and other similar firms.

As AI continues to evolve at a rapid rate, this risk is not theoretical - it is already materialising in two distinct ways:

1. **Substitution in media:** Much of the initial revenue at risk is in areas currently dedicated to providing music for audio-visual productions and computer games (such as specially commissioned scores and production music). This sub-sector faces an immediate existential threat from generative AI models trained without authorisation on their past works to compete directly against them.

2. **Dilution in streaming:** The proliferation of AI music on traditional streaming services risks diluting royalties for human creators and enabling bad actors to fraudulently derive revenue. The streaming platform Deezer reports that it is now receiving over 60,000 fully-AI tracks every day, representing 39% of their total uploads. In 2025 alone, the platform

detected over 13.4 million of these tracks. Crucially, this content is now cannibalising legitimate market share. AI tracks currently generate between 1-3% of all streams on the platform—a significant jump from just 0.5% in September 2024. Most concerningly, Deezer reports this rise is driven by bad actors and that up to 85% of these AI streams were fraudulent (played by bots to siphon royalties away from human creators). Concerningly, as the scale and power of AI increases, so will its capacity to generate more tracks leaving this number as simply a starting point.

This represents a real and imminent threat to the industry's foundation. As the This Is Music 2025 report notes, songwriters create the building blocks on which the wider £8 billion music industry relies.

If AI models are permitted to mine these building blocks without any requirement to secure a licence, the value derived from British music will be transferred directly to AI platforms for nothing. This will leave the rights holders who invest in that foundation with no equitable economic return and—crucially—no ability to object to their work being used to build the very products that seek to replace them.

We are at the very start of seeing, let alone understanding, the impact that AI is going to have on music publishing, our members and the wider creative community. As such, it is incredibly challenging to quantify its economic or creative impact.

Given that AI is clearly in the infancy of its development, we implore the government to explore a range of methods when trying to understand what this means in the longer term.

1. Strategic context and sector applicability

The MPA welcomes the Competition and Markets Authority's (CMA) intervention regarding Google's Strategic Market Status (SMS). We broadly support the CMA's finding that the power imbalance between the SMS firm and content creators has created a situation requiring regulatory correction.

The Secretary of State for Science, Innovation and Technology recently characterised the current technological shift as an "Industrial Revolution" that will transform the economy over the next decade. The MPA supports this vision. Our members are already active adopters of assistive AI, utilising platforms such as Studio VPro, Cyanite, and Musiio to manage copyright data at scale. We are pro-innovation, but for the AI sector to be sustainable, it must not undermine the upstream investment in the creative assets upon which it relies.

Distinction between Referral models and Consumption models

The Consultation correctly identifies that news publishers rely on search engines for referral traffic. However, music publishers operate on a consumption model. The value of a musical work is not realised through a 'click' but through the consumption of the work itself.

If a Generative AI (GenAI) output displays a copyrighted lyric or composition, the user's informational need is satisfied within the search interface. This constitutes a substitution of the primary product. Consequently, remedies designed to optimise Click-Through Rates (CTR) address the wrong economic metric for our sector.

2. Assessment of proportionality and economic impact

The MPA challenges the provisional decision to reject Crawler Separation (separating Search indexing from AI scraping) on proportionality grounds. The Consultation deems the ~£150 million compliance cost to Google 'onerous.'

This analysis underestimates the counterfactual harm. As detailed in our introduction, the £408 million annual revenue risk to our sector vastly outweighs the compliance cost to the SMS firm. While we recognise the legislative requirement for measures to be delivered in a certain way, the impact here is clear and we urge the government to recognise this discrepancy.

Specific risk to Production Music

The harm is not evenly distributed. The Sync and Production Music sector—which creates background music for TV, gaming, and advertising—faces an immediate existential threat. The *CISAC/PMP* study projects at least a 60% revenue decline for this sub-sector as AI models offer royalty-free substitutes. This will devastate the grassroots of the UK composing economy, which supports thousands of independent creators. Again, this figure is a starting point and the long-term impact of AI on production music will likely only increase.

Evidence of market distortion

This is supported by recent market data. As noted above, Deezer reports receiving over 60,000 AI-generated tracks daily (39% of total uploads). Furthermore, AI content now accounts for 1-3% of total streams, with up to 85% of those streams identified as fraudulent.

By rejecting Crawler Separation, the CMA is effectively prioritising a compliance saving for the SMS firm over the protection of a foundational UK industry facing measurable dilution.

3. Technical limitations of website-level opt-out mechanisms

The Consultation proposes requiring Google to provide 'effective controls' (such as Google-Extended) to allow publishers to opt out. While technically feasible for news publishers hosting content on proprietary domains, this is structurally unsuitable for the music sector.

We believe that proposing domain-level opt-outs is not only a technical failing but constitutes the formal imposition of an Unfair Trading Condition under Chapter II of the Competition Act 1998. Furthermore, it directly violates the 'Fair Dealing' objective of the SMS Conduct Requirements established under the Digital Markets, Competition and Consumers (DMCC) Act. Given the Google's sophisticated understanding of the internet's architecture, it is fully aware that dispersed rights holders cannot control third-party domains. By forcing a structurally unworkable compliance mechanism – that is essentially a 'take it or leave it' approach onto upstream suppliers, Google is being allowed to exploit its asymmetric market power and abuse its dominant gateway position to bypass fair commercial licensing.

Dispersed content distribution

The CMA's current definition of a 'Publisher' seems to presume website ownership. Crucially, thought, rights holders (songwriters/publishers) do not control or administer every website or online service where their content appears. A songwriter and/or a music publisher cannot edit the robots.txt file of a third-party lyric site, music service or video platform.

Aside from the lack of direct control over licensed third party websites and services, it is important to note that the internet is also an environment where copyrights owned and controlled by songwriters and music publishers are available at scale on an entirely unlicensed basis without any prospect of any input into any mechanisms used by unlicensed services to exercise any control over the use of the content they are making available by Google or other companies seeking to scrape that content for AI training. Therefore, an opt-out mechanism based on domain controls disenfranchises the actual copyright owner. If a lyric aggregator chooses not to opt out (to preserve traffic), the songwriter's work is scraped without their consent.

Even for news publishers, content is rarely confined to a single proprietary domain. News stories, extracts, and assets are distributed across the internet via social media, aggregators, and syndication partners. If a publisher can only control scraping on their own website, they have no effective means to prevent AI models from accessing their content once it appears on third-party platforms.

A regulatory remedy that relies on robots.txt or similar domain-level signals is therefore porous by design, as it fails to capture the reality of how content is distributed and consumed online.

This issue is even more acute for the music sector. Musical works are distributed across a highly fragmented ecosystem of third-party platforms (DSPs, lyric aggregators like Genius, social media). A technical opt-out signal on a music publisher's corporate website does not propagate to these third-party platforms.

Crucially, a songwriter cannot edit the robots.txt file of a third-party lyric site. Therefore, an opt-out mechanism based on domain controls disenfranchises the actual copyright owner. If a lyric aggregator chooses not to opt out (to preserve traffic), the songwriter's work is scraped without their consent.

Metadata stripping

Furthermore, metadata stripping by downstream platforms remains a significant technical issue; for example, social platforms routinely strip EXIF data upon upload. This removes the opt-out tags the CMA proposes to rely on and, as content travels, markers to stop scraping can be lost.

Ultimately, the proposed transparency metrics fail to address the severe information asymmetry between Google and content creators. By refusing to disclose verifiable ingestion or training logs, Google prevents rights holders from assessing the true commercial use of their assets. This prevents the formation of a functional, competitive licensing market, creating a situation that the CMA's conduct requirements are explicitly meant to resolve.

We also note a double standard in the market. While AI developers argue that mining data is 'fair game,' they explicitly prohibit the same activity on their own platforms. OpenAI's own Terms of Use prohibit users from 'scraping' their output to train competing models. The government should not mandate that creators accept scraping practices that AI companies themselves prohibit. As such, we suggest the regulatory framework should not rely on opt-out mechanisms. Instead, licensing must be the default position.

4. Transparency obligations regarding training data

The MPA considers that the transparency metrics proposed in the Consultation—specifically 'Search Impressions' and Click-Through Rates (CTR)—are technically unsuited for monitoring Text and Data Mining (TDM) activity.

Exclusion of automated traffic from standard reporting

The CMA proposes using Google Search Console as the primary transparency tool for publishers. However, this tool is engineered to track *human* user behaviour, not *machine* data extraction.

Google's own technical documentation for Search Console confirms that the platform systematically filters out 'crawlers' and 'robotic activity' to preserve data integrity for advertisers. Consequently, the specific bot traffic used to scrape content for AI training is excluded from the very reporting tool proposed as a safeguard. Relying on this metric creates a regulatory blind spot where the extraction of rights-protected content remains unreported to the rights holder.

Structural shift from Referral-based indexing to content Substitution

Focusing transparency obligations solely on *output* metrics (such as Search Impressions) fails to capture the primary economic risk: the transition of Google from an intermediary to a direct competitor.

It is critical to distinguish between *indexing* content (guiding a user to a source) and *synthesising* content (replacing the need for the source):

- **Referral (Search):** In a traditional search interaction, the engine acts as an intermediary. If a user searches for lyrics, the engine provides a link, and the user clicks through to the publisher's domain where the consumption event is monetised.

- **Substitution (AI):** In a GenAI interaction (e.g., AI overviews), the model generates the lyric or answer directly within the interface. The user's informational need is fully satisfied without them ever visiting the publisher's source.

This creates an interaction where the platform captures 100% of the user engagement while the rights holder receives 0% of the benefit. Consequently, measuring CTR is irrelevant in a system designed to keep the user on the search results page.

A low CTR in this context does not indicate low relevance—it indicates high substitution, confirming that the AI model has rendered the original source commercially redundant.

Evidence of memorisation and regulatory arbitrage

Recent legal proceedings have demonstrated that these risks are not theoretical but are currently occurring at scale.

- **Memorisation** - *Concord Music Group v Anthropic***:** Evidence presented in 2024 demonstrated that Large Language Models (LLMs) can retain and reproduce substantial portions of copyrighted lyrics verbatim. This confirms that the ingestion process results in the creation of unauthorised copies within the model's weights, allowing the AI to serve as a market substitute even if it does not link to the source. Another example of a similar situation exists in *GEMA v. OpenAI (Munich Regional Court).*

- **Data Laundering** - *Kneschke v LAION:* The recent German case of *Kneschke v LAION* highlighted a structural loophole where commercial AI developers use non-profit research datasets to bypass licensing requirements. This practice effectively washes data of its copyright obligations. Without transparency regarding the *source* of the data (the input), this form of regulatory arbitrage is impossible to detect.

The MPA suggests that, to mitigate these significant risks, and ensure effective oversight, transparency requirements must extend to the input of the model through an obligation on Google, as a firm with SMS, to disclose auditable training and ingestion logs including specific records of the URLs and domains scraped for model training. Additionally, we reject any reporting framework that relies solely on search console data, as this conflates indexing for search with scraping for training.

5. Feasibility of licensing frameworks

The MPA rejects the suggestion that licensing for AI training is technically unachievable.

It is not unusual for the tech sector to claim that a requirement to license the use of copyright would represent an obstacle to future developments that cannot reasonably be cleared. Hyperbolic arguments that requirements to secure licenses would "break the internet" were made in the context of the introduction of Article 17 of the European Copyright Directive with some entities going so far as to block access to their services in some European jurisdictions to put pressure on regulators not to implement the proposed legislative changes.

In practice the passage of Article 17 did not result in the death of the internet or the end of the availability of user generated content services – those services continued to thrive and have done so with the benefit of licenses that ensure that the music publishers and the songwriters they represent participate in the value created when their copyrights are used. In the context of Google in particular, the company already has a well-developed and scalable licensing infrastructure in place to power services such as You Tube.

The market has repeatedly demonstrated that voluntary licensing is viable for new and emerging complex uses of technology. Rights holders have already been successfully concluding licensing agreements with AI developers (e.g. Sony Music Publishing with Klay Vision; Universal Music Publishing with Udio; Warner Chappell with Suno and Kobalt with ElevenLabs). These agreements prove that a market-based solution is emerging.

By codifying an opt-out regime, the regulation risks undermining these voluntary solutions, validating a model where rights holders must actively withhold content rather than one where developers must secure permission.